The Danish National Child Health Register

Mikkel Porsborg Andersen,1,2 Rikke Wiingreen,3 Talip E Eroglu,4,5 Helle Collatz Christensen,6 Laura Bech Polcwiartek,7,8 Stig Nikolaj Fasmer Blomberg,6,9 Kristian Kragholm10,11 Christian Torp-Pedersen10,11, Kathrine Kold Sørensen1

1Department of Cardiology, Nordsjællands Hospital, Hillerød, Denmark; 2The Prehospital Center, Region Zealand, Denmark; 3Department of Pediatrics, Nordsjællands Hospital, Hillerød, Denmark; 4Department of Cardiology, Herlev-Gentofte Hospital, Copenhagen, Denmark; 5Department of Experimental and Clinical Cardiology, Heart Centre, Amsterdam Cardiovascular Sciences, Amsterdam UMC, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands; 6Copenhagen Emergency Medical Services, Copenhagen, Denmark; 7Department of Pediatrics, Randers Regional Hospital, Randers, Denmark; 8Department of Pediatrics, Aalborg University Hospital, Aalborg, Denmark; 9Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 10Department of Cardiology, Aalborg University Hospital, Aalborg, Denmark; 11Unit of Clinical Biostatistics and Epidemiology, Aalborg University Hospital, Aalborg, Denmark; 12Department of Public Health, University of Copenhagen, Copenhagen, Denmark.

Correspondence: Mikkel Porsborg Andersen, Department of Cardiology, Nordsjællands Hospital, Dyrehavevej 29, Hillerød, 3400, Denmark, Tel +45 92 43 22 99, Email mikkel.porsborg.andersen@regionh.dk

Aim of the Database: The aim of the National Child Health Registry is to provide comprehensive insight into children’s health and growth on a national scale by continuously monitoring the health status of Danish children. Through this effort, the registry assists the health authorities in prioritizing preventive efforts to promote better child health outcomes.

Study Population: The registry includes all Danish children, however, incomplete coverage persists.

Main Variables: The National Child Health Registry contains information on exposure to secondhand smoking, breastfeeding duration, and anthropometric measurements through childhood. The information in the registry is divided into three datasets: Smoking, Breastfeeding, and Measurements. Beside specific information on the three topics, all datasets include information on CPR-number, date of birth, sex, municipality, and region of residence.

Database Status: The National Child Health Registry was established in 2009 and contains health information on children from all Danish municipalities, collected through routinely performed health examinations conducted by general practitioners and health nurses.

Conclusion: The National Child Health Register is an asset to epidemiological and health research with nationwide information on children’s health and growth in Denmark. Due to the unique Danish Civil Registration System, it is possible to link data from the National Child Health Register to information from several other national health and social registers which enables longitudinal unambiguous follow-up.

Keywords: child health, children’s database, register-based research, Danish register, epidemiology

Introduction

The National Child Health Registry, sometimes referred to as the Children’s database,1-3 was established in 2009 and contains information on children’s health by collecting information from routinely performed health examinations conducted by general practitioners, health nurses, and municipalities.4 The purpose of the registry is to gain insight into children’s health and growth on a national scale by following the health status of Danish children (ie, aged 0–17 years) which can aid the health authorities to prioritize preventive efforts.4 The insights also enable the 98 Danish municipalities to compare themselves with each other, which may further help to measure the effect of preventive efforts. Also, the registry includes research possibilities, especially from a pediatric perspective where investigating factors that may affect childhood health and development may play a crucial part in our understanding of various diseases and conditions. In addition, several studies have stated that establishing preventive measures to avoid chronic diseases is easier and more effective during childhood and adolescence than in adulthood.5-7 Thus, it is important to understand how factors such as passive smoking, breastfeeding, or body-composition affect children’s early years of life as these may...
affect long-term health and development. However, nationwide registry data on long-term follow-up for children are limited and difficult to gain access to. Considering this, the National Child Health Registry has a relevant public health implication as delineating factors affecting children’s health may result in behavioral benefits. Although most of the Danish registries have been described previously,8–19 a detailed description of the Danish National Child Health Registry has not yet been presented. The purpose of this paper is therefore to provide a thorough description of variables and information available in the Danish National Child Health Registry as well as give insights into the strengths and limitations of the registry.

Setting
Denmark has a long tradition of collecting healthcare and administrative information on all Danish residents with a yearly population coverage of approximately 5.8 million individuals, with some registries starting already before the 19th century.15 Since then, the data collection has extended, and with the introduction of the Danish Civil Registration System in 1968, which includes the unique civil personal registration number (CPR-number), the opportunity to link information from different registries has increased dramatically. The CPR-number is assigned to all Danish residents at birth or immigration and is included in all Danish nationwide registries, thereby enabling linkage of information from the different registries on an individual level.13 This possibility and the variety of economic, social, and healthcare information available from the registries on the complete Danish population is probably the reason why Denmark is often referred to as one large cohort. A deeper insight into the Danish setting has been described elsewhere.15,20,21 The Danish registers therefore provide unique opportunities for conducting nationwide register-based studies within the fields of epidemiology, health, economics, and social sciences, among others, where the nationwide design minimizes the risk of selection bias as all age, ethnic, and socioeconomic groups are represented. The nationwide register-based design likewise limits the risk of information bias due to the use of high-quality data collected for administrative purposes enhancing information validity as well as limiting loss-to-follow. In this relation, knowledge of information available through the different registries is essential in choosing a research question and study design. Also, of notice, some Danish registries might hold similar or related information, but these will often differ in regard to the context, time, and place the information was collected, which likewise can be utilized in the study design, especially if the aim of the research question is to assess time-varying exposure, confounding, or mediators.

Access to the Registries
In Denmark, there are two main organizations where researchers have the possibility of getting access to the Danish registries. The two organizations are the Danish Health Data Authority which is responsible for the Danish health registries, and Statistics Denmark which is responsible for the Danish social registries. In relation to Danish legislation, specifically The Data Protection Act, it is possible to get access and process information from the registries if the objective is to perform statistics and scientific research of significant societal importance if processing of data is necessary for the execution of the investigations.22 Neither ethical approval nor informed consent is required if the previously mentioned objective is fulfilled.22 Register-based studies must be approved by the researcher’s data responsible institute in accordance with the General Data Protection Regulation (GDPR).23

Information in the National Child Health Registry
The National Child Health Registry contains information on exposure to secondhand smoking, breastfeeding duration, and anthropometric measurements. The information in the registry is divided into three datasets which throughout this paper will be referred to separately as Smoking, Breastfeeding, and Measurements.
All datasets include variables on background information including the CPR-number, date of birth, sex, municipality of residence, and region of residence. A full description of all available variables in the datasets is included in the supplementary material (Table S1). Figure 1 shows how many out of the total number of live births by year of birth, derived from the Civil Registration System, there is information on in the three datasets. Although the registry was established in 2009, it only became mandatory for all Danish municipalities to report data to the registry in December 2011. As a result, the registry’s validity has been recognized from 2012 onward,24 which is reflected in the
coverage (Figure 1). The exact number of individuals in each birth cohort and coverage percentages are included in the supplementary material (Table S2). The Measurement dataset has a high coverage for all shown birth years, whereas the Smoking dataset does not cover 80% until 2013 and the Breastfeeding dataset never reaches a coverage above 70%. However, the Breastfeeding dataset, as per the definition of the registry, will never reach full coverage since not all mothers breastfeed. The data available for this study include the entirety of 2019. All children’s health examination results are reported to The National Child Health Registry through the Danish Health Data Authority’s electronic reporting system called SEI2. It is mandatory for all the municipalities’ health nurses to report information from the examinations on height, weight, duration of full breastfeeding, and exposure to tobacco smoking in the home at birth to SEI2, while it is voluntary for the general practitioners to report to the SEI2 system.

Relevant to all three datasets, all parents in Denmark are offered at least five free home visits by a child health nurse throughout the child’s first year with the purpose of providing guidance, monitoring the health and well-being of the child, and collecting standardized information.

Smoking
Exposure to secondhand smoke in children increases the risk of sudden infant death syndrome, pneumonia, and severe asthma symptoms. Moreover, it results in thousands of avoidable hospitalizations. The Smoking dataset contains information regarding children’s exposure to tobacco smoking at home in the first year of the children’s lives. Exposure to smoking has since 2012 been registered by child health nurses during home visits. However, the dataset includes data entries for children born much earlier, for example, there is information on exposure to smoking for 20 children born in 1980 and this number only increases in the subsequent years, which could be due to pilot projects or errors. Registered exposure to tobacco smoke including E-cigarettes implies that at least one resident in the child’s home smokes inside the home at least once a week. There is no information on exposure to tobacco smoke from persons outside the home, eg, grandparents. The dataset includes one entry per child, if multiple entries existed, only information from the latest date was included. Beyond the variables including background information, the dataset includes a variable with the numeric code denoting the exposure (variable: “PASSIVRYGNINGKODE”) and a variable with the code-specific text (variable: “PASSIVRYGNINGNAVN”), indicating exposure to secondhand smoking. Exposure to tobacco smoke at home is included in three categories (0, 1, and 2). Value “0” indicates exposure to tobacco smoke at home (ie, passive smoking),
“1” indicates not being exposed to tobacco smoke at home (ie, no passive smoking), while “2” indicates unknown exposure to tobacco at home. As of 31 December 2019, the smoking dataset included 882,962 individuals, of which 57.8% were classified as unknown, 2.1% as exposed for passive smoking, and 40.1% as not exposed. Figure 2 shows the proportion of children in the Smoking dataset registered as exposed, unexposed, or unknown to secondhand smoking for each birth year from 2000 to 2019. The high percentage of unknowns should be interpreted with respect to the low coverage of information regarding exposure to secondhand smoking in the Smoking dataset, as shown in Figure 1.

Breastfeeding
Breastfeeding provides health benefits for both the mother and the child, and information regarding the duration of exclusive breastfeeding is important when evaluating children’s health. The information on exclusive breastfeeding is collected by the child health nurses at routine home visits in the first year of life. In the breastfeeding dataset, the data coverage is reliable from 2012 and onwards but entries also include children born much earlier. In the dataset, exclusive breastfeeding is registered if the infant is fed only with breast milk in addition to water and/or a maximum of one bottle of infant formula per week after hospital discharge. The duration time is registered in completed days, calculated by subtracting the date of birth from the date registered as the stop date for exclusive breastfeeding. The dataset includes one entry per child, if multiple entries existed, only information from the latest date was included. Entries stating durations above 365 days are excluded as these are deemed unrealistic by the Danish Health Authority. It is prudent to note that current guidelines state that if an infant does not begin exclusive breastfeeding, there will be no record of cessation of exclusive breastfeeding and consequently, no data for that infant in the dataset. However, some infants have a cessation date equal to their date of birth, which indicates that they did not initiate breastfeeding which may be due to local routines. Therefore, only children who have initiated breastfeeding can be found in the dataset, according to the definition, but caution is advised. Importantly, the absence of a record cannot be assumed to indicate a lack of breastfeeding, as missing data for other reasons is expected. The dataset holds the following central information beyond the background information described in the introduction: stop date for exclusive breastfeeding (variable: “DATOSTOPPET”) and duration time (variable: “VARIGHED_DAGE”). As of 31 December 2019, the breastfeeding dataset included 341,375 individuals and the breastfeeding duration median was 126 days with an interquartile range of 49–165 days. Figure 3 depicts the percentage of children born in the respective birth years that are registered as having...
been breastfed for less than one month, one to four months, four to six months, or more than six months, respectively. The numbers from the earlier birth years should be interpreted with respect to the lower coverage (Figure 1).

**Measurements**

The aim of the measurement dataset is to monitor the prevalence/frequency of Danish children with underweight, normal weight, overweight, and obesity throughout childhood. The measurements originate from three sources based on the age of the child when it is performed. From age 0–1 years, the measurements are performed by child health nurses at routine home visits. From age 1–5 years the measurements are performed and reported by the general practitioner, or child health nurse in some cases, in relation to routine preventive child health examinations. The measurements from ages 5–17 years are performed by the school health nurse in relation to the regular health checkups in the school setting which are offered as a preventive measure for all children living in Denmark. It is intended that measurements are performed at least once at the beginning of compulsory education when the child is 5–6 years old, once in middle school when the child is between 7 and 12 years old, and finally once in the last part of compulsory education where the child is 13–16 years old. Each child can thus have several measurements at different time points for the same variable.

The dataset holds the following central information beyond the background information described in the introduction: The code and name identifying the school level (variable: “SKOLEGRUPPEKODE” and variable: “SKOLEGRUPPENAVN”), information on weight (variable: “VAEGT_GRAM”), height (variable: “HOEJDE_CM”), and date of the measurement (variable: “DATE”). For the relevant age groups, the Body mass index (BMI)-category is also stated as a code and text (variable: “BMIKATEGORIKODE” and variable: “BMIKATEGORINAVN”). Height is registered in centimeters, with no decimals, rounded for each 0.5 centimeters. Children under the age of two years are measured with a measuring tape lying flat on their back whereas children at the age of two years or older are measured barefoot standing by, but not touching, a vertical edge with gathered legs. Children under the age of one year are weighed without clothes and diapers whereas older children are weighed barefoot, without outerwear, and with empty pockets, registered in grams. The BMI-category is applicable for ages two years and older and is based upon age- and sex-standardized BMI-curves developed by The International Obesity Task Force (IOTF). The BMI curves are based on height and weight data, including almost 20,000 boys and girls aged 2–18 years, originating from six representative countries (Brazil, UK, Hong Kong, The Netherlands, Singapore, and USA), measured in the years 1963–1993. Accordingly, the cutoffs have been presented and the curves have been widely used in the scientific literature which...
enables international comparisons of children’s BMI.\textsuperscript{31,34} Figure 4 shows the distribution of children registered as underweight, normal weight, overweight, and obese, shown for each birth year and the four age groups. If a child had multiple measurements taken within the same age group, only the first measurement within each age group was included in the figure. In the Measurement dataset, as of 31 December 2019, 85% of the children had more than one measurement registered. The median number of measurements per individual was 4, with an interquartile range of 2–6. In total, the measurement dataset included 6,709,833 observations and 1,470,655 individuals. The median age at measurements was 0.8 years, with an interquartile range of 0.2–7.9 years, the median weight was 9800 grams, with an interquartile range of 5500–28,400 grams, and the median height was 75 cm, with an interquartile range of 60–131 cm, the latter metrics being highly affected by the repeated measurements during early life.

Discussion

The main strength of the Danish National Child Health Registry resides in its population-based design in which each case is registered for purposes other than research, thereby minimizing the risk of selection bias, and rendering the findings from this registry representative. For all the three datasets, the information is collected as a part of the Danish public health system which is free and accessible for all. The information is reported in a uniform manner through electronic reporting systems, thereby mitigating the possibility that data is recorded by different methods. Consequently, the registry offers valuable opportunities for researching early-life exposures and long-term health, potentially influencing preventive strategies. Nevertheless, the Danish National Child Health Registry also has several limitations. A limitation of the data is its lack of completeness, particularly before it became mandatory (ie, before 2012), and the dwindling completeness after 2017 (as shown in Figure 1). This has been a persistent challenge, and the Danish Health Data Authority reports that some municipalities have periodically lower reporting frequencies than expected, possibly due to technical difficulties. However, these challenges are continuously being addressed in an effort to improve coverage.\textsuperscript{28} Post-registration is possible because the registry is regularly updated, and these updates can include information dating back several years. In general, the observed underreporting could not be attributed to any discernable structural distinctions, such as regional variations in reporting. Possible contributing factors could encompass ethnic background, socio-economic disparities, or errors, yet these factors remain speculative. Caution is warranted when considering the use of missing data handling methods, such as imputation, as missing at random cannot be assumed. Another limitation is the fact that reporting is not mandatory for
general practitioners and that the scope of this missingness is unknown. When using these datasets for new studies, it is thus important to consider the potential for selection bias due to incomplete coverage. Furthermore, a limitation is that the registry does not include information on the intensity of children’s exposure to tobacco smoking at home. Likewise, the registry only includes information on breastfeeding in children in whom breastfeeding was actually initiated. There are also some inconsistent and potentially flawed data entries in the datasets. Researchers working with the dataset should therefore address these errors carefully. Examples of flawed data include the entry of 0 or negative values in the Breastfeeding dataset, which may be interpreted as “not breastfed” but the meaning is not unambiguously defined since children who were not breastfed should have been omitted from this dataset. Conversely, other entries seem implausibly high, given that the definition is “exclusively breastfed”. Similarly, some entries in the Measurement dataset result in physiologically implausible BMI calculations, either too low or too high. Some of these implausible entries appear to be entry errors where the exact same value has been entered both as height and weight. Plausible typing errors may also exist, for example for some height and weight entries where it seems that an extra digit was added or missing a digit entirely. Furthermore, there are duplicates in all three datasets. The above-mentioned outliers and erroneous data entries should be carefully and transparently evaluated to ensure the accuracy of the information in the registry. In sum, the variation in data entries could affect the validity and therefore future validation studies of the data collected in the registry are warranted. However, the registry gives an important insight into early life exposures which can be combined with information from multiple other Danish registries, providing opportunities for valuable research.

Conclusion
The National Child Health Register is a valuable asset in epidemiological and health research with nationwide information on children’s health and growth in Denmark since 2012. Due to the unique Danish Civil Registration System, it is possible to link data from the National Child Health Register with information from several other national health and social registers, which enables longitudinal unambiguous follow-up. As the validity of some datasets varies, caution is needed when conducting research with The National Child Health Registry.

Ethics
This type of study does not require ethical approval or informed consent as stated in The Danish Data Protection Act in that the study is conducted for the sole purpose of statistics and scientific research. However, approval to use the data and conduct the study was granted by the data responsible institute in the Capital Region of Denmark (approval number P-2019-280) in accordance with the General Data Protection Regulation (GDPR).

Disclosure
Prof. Dr. Christian Torp-Pedersen reports grants from Bayer, grants from Novo Nordisk, outside the submitted work. The authors report no other conflicts of interest in this work.

References